

A. Examine window of typed SNPs

Reference haplotypes aligned with genotypes at the typed SNPs (untyped SNPs not needed)



One sample's unphased genotypes. One marker is missing due to genotyping error

Reference haplotypes

h_1	1	1	0	0	1	0
h_2	1	1	0	0	1	0
h_3	1	1	0	0	1	0
h_4	1	0	1	0	0	1
h_5	1	0	1	0	0	1
h_6	1	0	1	0	0	1
h_7	0	0	1	0	1	0
h_8	0	0	1	0	1	0

Target genotype

x_1	2	1	0	0	?	1
-------	---	---	---	---	---	---

B. Find optimal haplotype pairs in each window

Find unique haplotypes

Unique reference haplotypes

h_1	1	1	0	0	1	0
h_4	1	0	1	0	0	1
h_7	0	0	1	0	1	0

$$\|x'_1 - h_1 - h_4\|^2 = 2$$

$$\|x'_1 - h_1 - h_7\|^2 = 2$$

$$\|x'_1 - h_4 - h_7\|^2 = 3$$

Initialize missing

x'_1	2	1	0	0	0	1
--------	---	---	---	---	---	---

Find h_1, h_4 using least squares

$$h_1 + h_4 \approx x_1$$

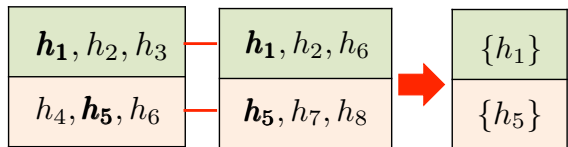
Extract matching haplotypes

strand 1: $\{h_1, h_2, h_3\}$
strand 2: $\{h_4, h_5, h_6\}$

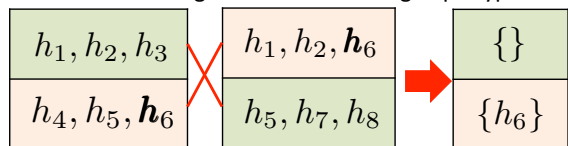
Candidate haplotypes

C. Connect neighbors in 1 of 2 ways

Parallel connection generates 2 surviving haplotypes:



A switch at window 2 generates 1 surviving haplotype:



D. Stitch window-by-window from left to right

Unphased haplotypes

h_1, h_2, h_3	h_1, h_2, h_6	h_1, h_3	h_4, h_5, h_8
h_4, h_5, h_6	h_5, h_7, h_8	h_2, h_5	h_2, h_6

Phased haplotypes

h_1	h_1	h_1	h_2, h_6
h_5	h_5	h_5	h_5

Breakpoint

Switch

Window 1

Window 2

Window 3

Window 4